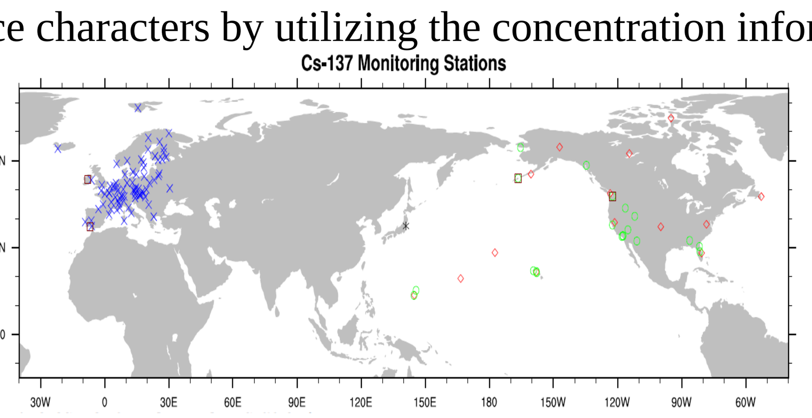
Data assimilation and inverse modeling with HYSPLIT Lagrangian dispersion model and satellite data - Applications to volcanic ash and wildfire smoke predictions

Tianfeng Chai[1,2], HyunCheol Kim[1,2], Ariel Stein[1], Fantine Ngan[1,2], Alice Crawford[1,2], Barbara Stunder[1], Michael J. Pavolonis[3]

1. NOAA Air Resources Laboratory, College Park, MD
2. Cooperative Institute for Climate and Satellites, University of Maryland, College Park, Maryland
3. NOAA Center for Satellite Applications and Research, Madison, WI

HYSPLIT Dispersion Model and its Inverse Modeling System

The HYSPLIT Lagrangian model developed at NOAA's Air Resources Laboratory has been widely used to study the atmospheric pollutant transport and dispersion in both forward and backward modes. It allows representations of the transported air masses with 3D particles, puffs, or a hybrid of the two. Applications include the simulation of atmospheric tracer release experiments, radionuclides, smoke originated from wild fires, volcanic ash, mercury, and wind-blown dust, etc. A HYSPLIT inverse modeling system has been developed to quantify the source characters by utilizing the concentration information.



Data source	Number of monitoring stations	Count of total samples
CTBT	14	417 (421)
EPAR	19 (20)	35 (39)
EURO	78 (80)	785 (797)
EXTR	4	59 (61)
Total	115 (118)	1296 (1318)

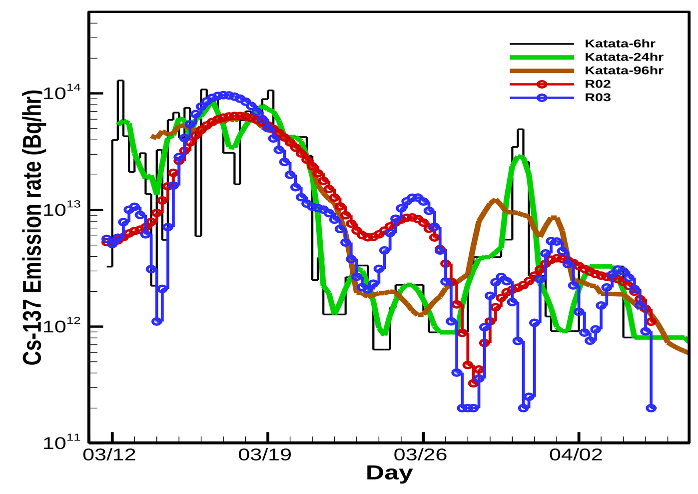


Figure 1. HYSPLIT inverse modeling application - Fukushima Nuclear Accident. Left: Distribution of Cs-137 monitoring stations and count of 24-hour concentrations; Right: comparison of HYSPLIT Cs-137 emission estimates and Katata et al. (2015) results. Smoothness parameters are 1.0 and 0.1 for R02 and R03, respectively. Details can be found in the following paper, Chai, T., R. R. Draxler, and A. Stein, *Atmospheric Environment*, 106, pp. 241-251, doi:10.1016/j.atmosenv.2015.01.070, 2015

HYSPLIT Inverse Modeling Methodology

In this top-down approach, the unknown emission terms are obtained via searching the emissions that would provide the best model predictions closely matching the observations by minimizing a cost function. For the applications shown here, the emission locations are mostly identified, the unknown emission rates and sometimes the release heights are left to be determined. The emission rates may vary significantly with time. Thus, the unknowns of the inverse problem are the emission rates q_{ikt} at each location *i*, at different height *k* and period *t*. The cost function *F* is defined as,

$$\mathcal{F} = \frac{1}{2}\sum_{t=1}^{T}\sum_{k=1}^{K}\sum_{i=1}^{I}\frac{(q_{ikt}-q_{ikt}^{b})^2}{\sigma_{ikt}^2} + \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{M}\frac{(c_{nm}^{h}-c_{nm}^{o})^2}{\epsilon_{nm}^2} + \mathcal{F}_{other} \quad (1)$$

where c^o_{nm} is the *m-th* observed concentration or mass loading at time period *n* and c^h_{nm} is the HYSPLIT counterpart. As shown in Equation (1), a background term is included to measure the deviation of the emission estimation from its first guess q^b_{ikt}. The background terms ensures that the problem is well-posed even when there are not enough observations available. The background error variances σ^2_{ikt} measure the uncertainties of q^b_{ikt}. The observational error variances ε^2_{nm} represent the uncertainties from both the model and observations as well as the representative errors. F_{other} refers to the other regularization terms that can be included in the cost function. The optimization problem can be solved using many minimization tools, such as L-BFGS-B package, to get the final optimal emission estimates.

Improving Volcanic Ash Predictions with HYSPLIT and Satellite Retrievals - 2008 Kasatochi Eruption

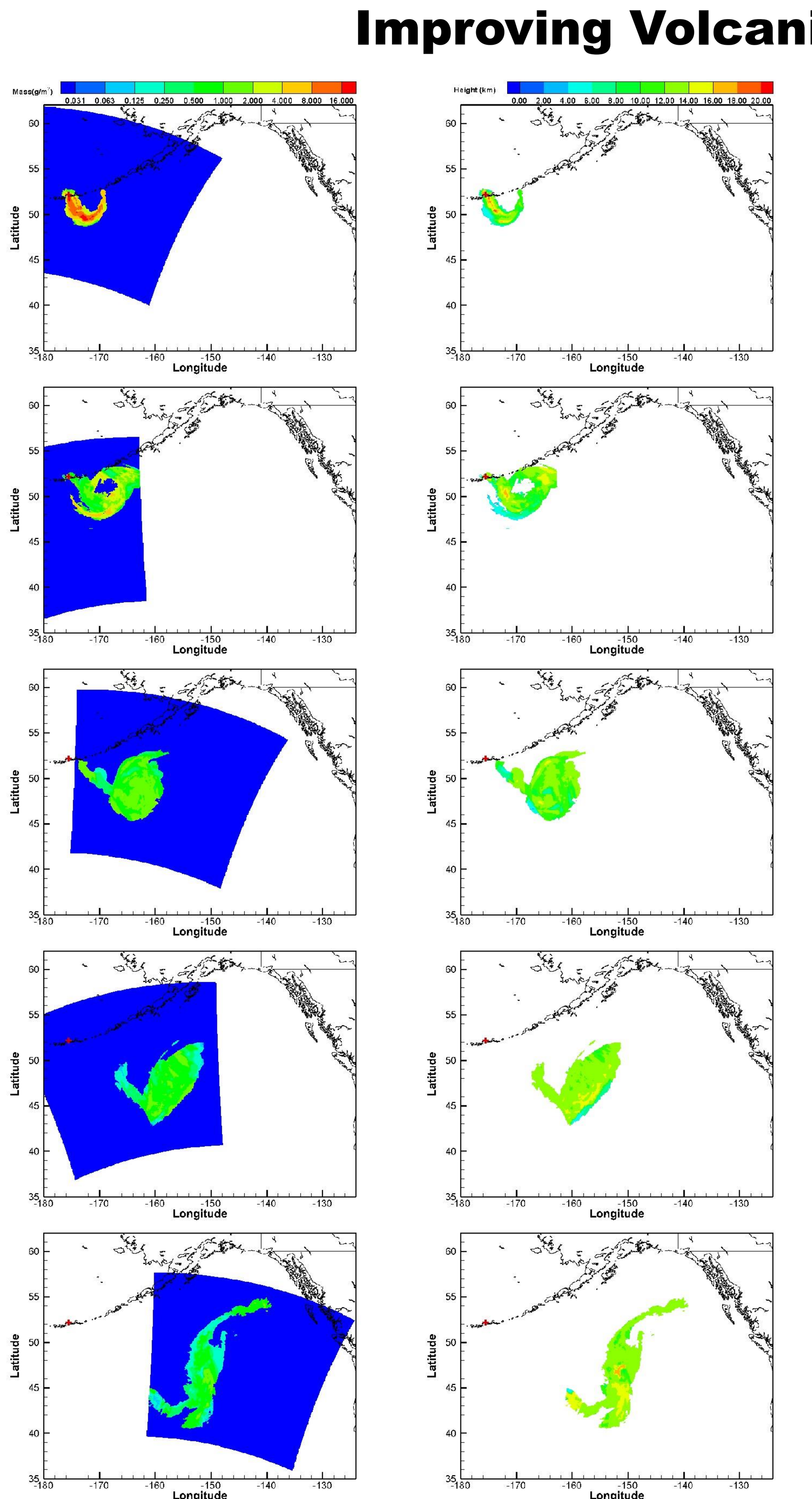


Figure 2. MODIS volcanic ash mass loadings (left) and ash plume top height (right) of the 2008 Kasatochi eruption in Aleutian Islands (shown with "+"). Observations (G1-G5) from top to bottom, 1340Z on Aug 8, 0050Z on 9 August, 1250Z on 9 August, 0000Z on 10 August, and 1150Z on 10 August, 2008.

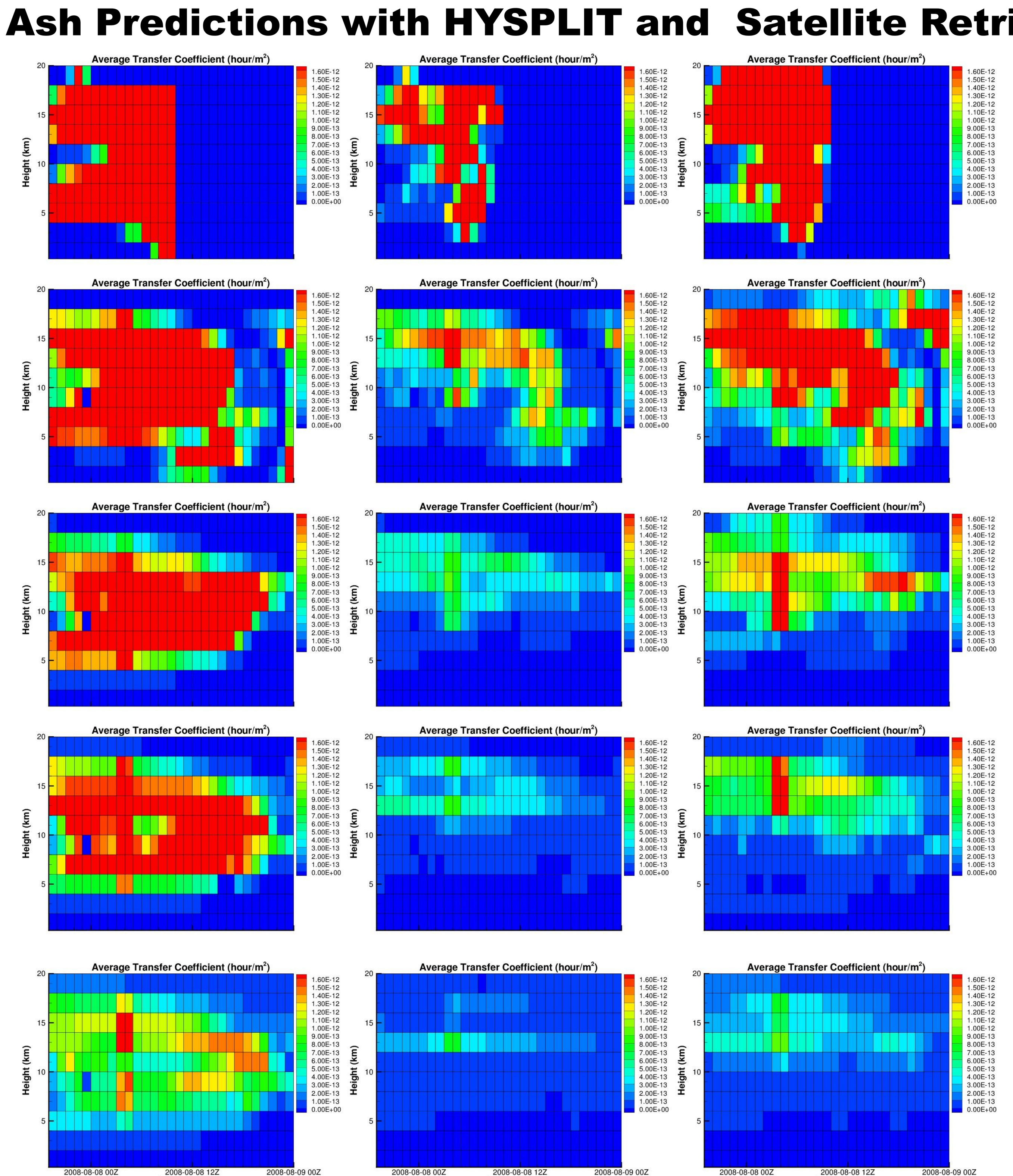


Figure 3. Averaged transfer coefficients (source-receptor sensitivities) with three different options in calculating model mass loadings (left: integrating from surface to cloud top; center: calculated for a single layer where the cloud top height resides; right: integrating over three layers centered at the cloud top layer). Rows 1-5 correspond to observations G1-5 (see Fig.2 caption for observation time).

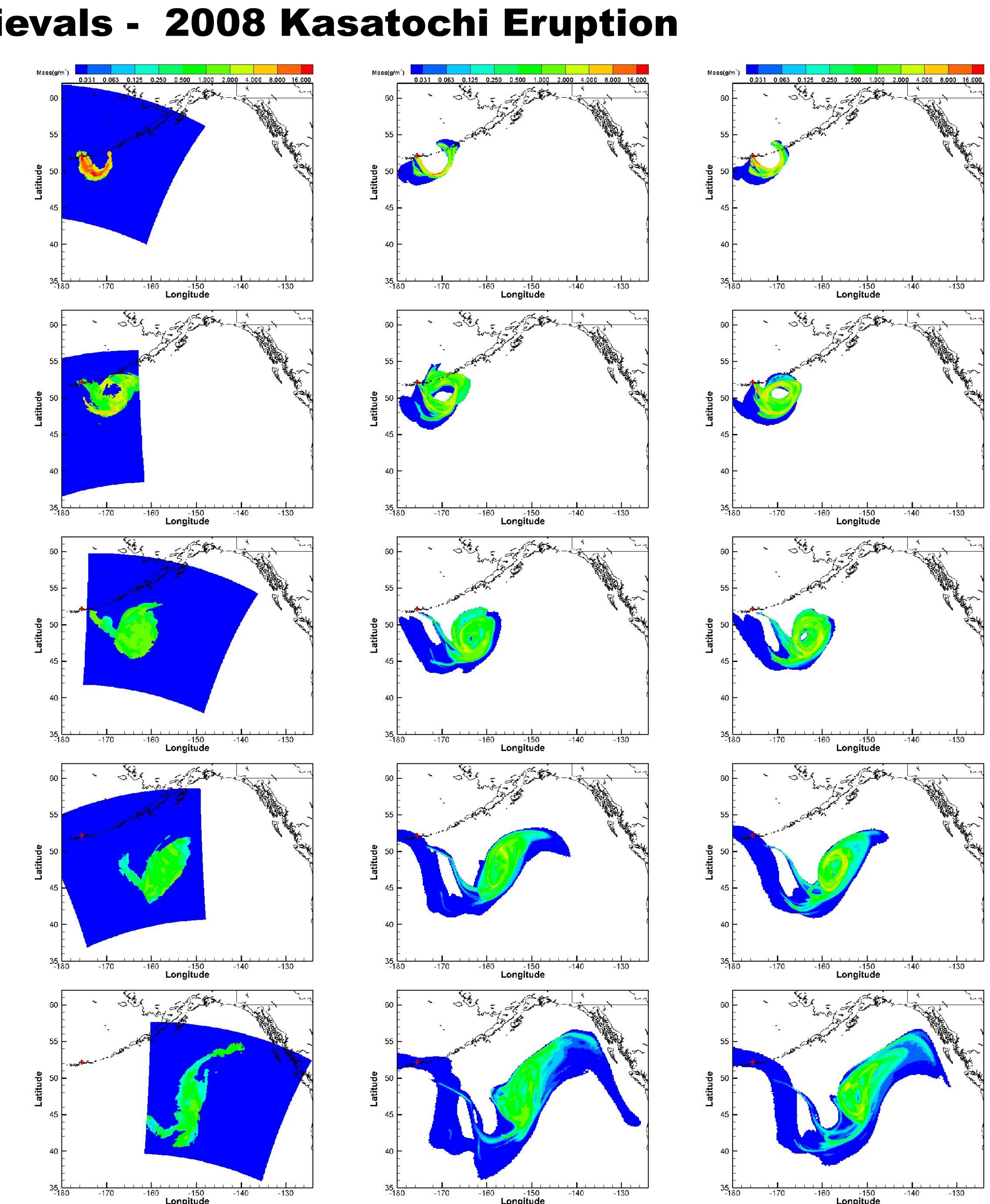


Figure 4. Volcanic ash mass loadings from MODIS (left) and HYSPLIT simulations with GDAS (center) and ECMWF (right), following their observation time from top to bottom (see Fig. 2 caption for detail).The ash release rates for the HYSPLIT simulations were obtained by assimilating granules G1,G2,and G3. In the inverse modeling, only ash pixels were used and the model mass loadings are calculated by integrating over three layers centered at the cloud top layer (option 3 in Fig. 3).

Wildfire Smoke

HYSPLIT is used for the operational smoke forecasts in support of the National Air Quality Forecast Capability (NAQFC). While most wild fire locations are well identified by the NOAA NESDIS Hazardous Mapping System (HMS), the current US Forest Service (USFS) BlueSky emission prediction may bring large uncertainties. This research aims to objectively and optimally estimate the wildfire smoke source strengths and their temporal variations based on NOAA NESDIS GOES Aerosol/Smoke products (GASP). Before applying real data, twin experiments are used to test the system with pseudo data.

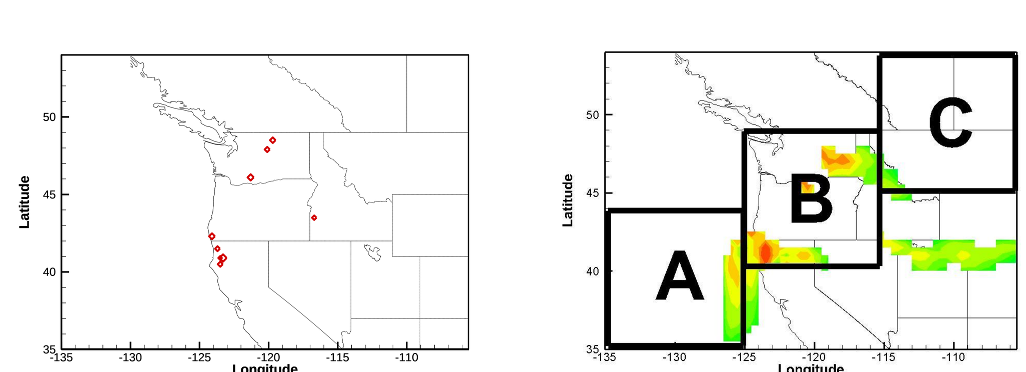


Figure 5. In the twin experiments, smoke releases at 9 fire locations (shown in the left panel) from 1500m or 2000m height at constant rates for two days from 6Z on August 17, 2015 are simulated by HYSPLIT. Hourly pseudo-observations of satellite mass loadings are generated based on the HYSPLIT results (right panel).

Inverse results
"Actual" sources

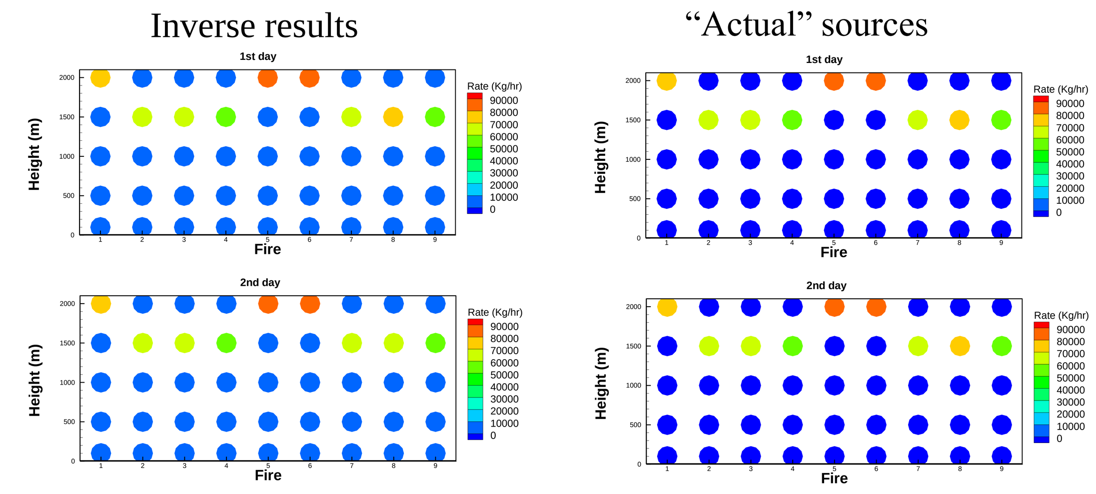


Figure 6. Comparison of the estimated smoke emission rates (left) at the nine locations (fire 1-9 as x-axis) for the two days (upper: 1st day; lower: 2nd day) and the "actual" sources (right) used in the twin experiment.

Table 1. Source term error statistics of the twin experiments

	Source term	MAE (kg/hr)	Normalized MAE	RMSE (kg/hr)	Normalized RMSE
Case 1 : Observations available at all 48 hours	Day 1	534.9	0.77%	841.4	1.21%
	Day 2	1760.5	2.53%	3332.5	4.78%
Case 2 : Observations available at last 24 hours	Day 1	1985.8	2.85%	3310.2	4.75%
	Day 2	1393.0	2.00%	2943.2	4.22%
Case 3: Observations available at last 24 hours; Region A in Fig 7 blocked.	Day 1	606.4	0.87%	1156.3	1.66%
	Day 2	301.2	0.43%	573.4	0.82%
Case 4: Observations available at last 24 hours Region B in Fig 7 blocked.	Day 1	23834.6	34.21%	32157.9	46.16%
	Day 2	66177.5	94.99%	78653.3	112.90%
Case 5: Observations available at last 24 hours Region C in Fig 7 blocked.	Day 1	3974.9	5.71%	8803.3	12.64%
	Day 2	3400.6	4.88%	10663.2	15.31%

Summary

- A HYSPLIT inverse system based on a 4D-Var approach has been built and successfully applied to estimate caesium-137 release during the Fukushima nuclear accident;
- The system is also used to solve the effective volcanic ash release rates as a function of time and height by assimilating satellite mass loadings and ash cloud top heights. Using the Kasatochi eruption in 2008 as an example, we show that the estimated releases help to improve the volcanic ash predictions with both GDAS and ECMWF meteorological fields.
- The feasibility of applying the method to objectively and optimally estimate wildfire smoke sources based on satellite observations of fire plumes is demonstrated using a set of twin experiments. They show promising results although missing observations in key regions could hinder the effectiveness.

Acknowledgement

This study was supported by NOAA award NA16OAR4590121 and NOAA grant NA09NES4400006 (Cooperative Institute for Climate and Satellites-CICS) at the NOAA Air Resources Laboratory in collaboration with the University of Maryland. Additional funding for this study was provided by the U.S. Department of State and the Federal Aviation Administration (FAA).

References

1. Source term estimation using air concentration measurements and a Lagrangian dispersion model – Experiments with pseudo and real cesium-137 observations from the Fukushima nuclear accident, by Chai, T., R. R. Draxler, and A. Stein, *Atmospheric Environment*, **106**, pp. 241-251, *doi:10.1016/j.atmosenv.2015.01.070*, 2015
2. Improving volcanic ash predictions with the HYSPLIT dispersion model by assimilating MODIS satellite retrievals, by Chai, T., A. Crawford, B. Stunder, M.J. Pavolonis, R. Draxler, and A. Stein,, *Atmos. Chem. Phys.*, **17**, pp. 2865-2879, *doi:10.5194/acp-17-2865-2017*, 2017

For questions and comments, please contact Tianfeng Chai (Email: Tianfeng.Chai@noaa.gov)